

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: The Scripps Research Institute
- (B) STREET: 10666 North Torrey Pines Road, Suite 220,  
Mail Drop TPC8
- (C) CITY: La Jolla
- (D) STATE: CA
- (E) COUNTRY: USA
- (F) POSTAL CODE (ZIP): 92037
- (G) TELEPHONE: 619-554-2937
- (H) TELEFAX: 619-554-6312

## (ii) TITLE OF INVENTION: SYNTHESIS OF PROTEINS BY NATIVE CHEMICAL LIGATION

## (iii) NUMBER OF SEQUENCES: 20

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: RatentIn Release #1.0, Version #1.25 (EPO)

## (v) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER: PCT/US
- (B) FILING DATE: 04-MAY-1995

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (ix) FEATURE:

- 43 -

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 5  
(D) OTHER INFORMATION: /label = COSH  
/note = "Wherein COSH is thioacid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Leu Tyr Arg Ala Gly  
1 5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Cys Arg Ala Glu Tyr Ser  
1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 5  
(D) OTHER INFORMATION: /label = COSBn  
/note = "Wherein COSBn is benzyl thioester."

09710633-110800

- 44 -

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Leu Tyr Arg Ala Gly  
1 5

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 5
- (D) OTHER INFORMATION: /label = X  
/note = "Wherein X is N-acetyl-cysteine-thioester."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Leu Tyr Arg Ala Gly  
1 5

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Leu Tyr Arg Ala Gly Cys Arg Ala Glu Tyr Ser  
1 5 10

00807T"EE90T250

(A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

/note = "Wherein/SCH<sub>2</sub>COOH is 2-thioacetic acid."

Leu Tyr Arg Ala Gly  
1 5

(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

```
/note = "Wherein COSH is thioacid."
```

/note = "Wherein Msc is 2-methyl-sulfonyl-ethyloxy-carbonyl."

[illegible]

Ala

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

**(ix) FEATURE:**

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 33  
(D) OTHER INFORMATION: /label= COSBn  
/note= "Wherein COSBn is benzyl thioester."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Ala

(2) INFORMATION FOR SEQ ID NO:9:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 39 amino acids  
(B) TYPE: amino acid

- 47 -

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Cys Ala Asn Thr Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu Leu  
1 5 10 15

Cys Leu Asp Pro Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys Phe  
20 25 30

Leu Lys Arg Ala Glu Asn Ser  
35

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(A) NAME/KEY: Modified-site  
(B) LOCATION: 72  
(D) OTHER INFORMATION: /label = SH4

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Ser Ala Lys Glu Leu Arg Cys Gln Cys Ile Lys Thr Tyr Ser Lys Pro  
1 5 10 15

Phe His Pro Lys Phe Ile Lys Glu Leu Arg Val Ile Glu Ser Gly Pro  
20 25 30

Ala Cys Ala Asn Thr Glu Ile Ile Val Lys Leu Ser Asp Gly Arg Glu  
35 40 45

008077-CE907260

- 48 -

Leu Cys Leu Asp Pro Lys Glu Asn Trp Val Gln Arg Val Val Glu Lys  
 50 55 60

Phe Leu Lys Arg Ala Glu Asn Ser  
 65 70

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 40
- (D) OTHER INFORMATION: /label = COSNB  
 /note = "Wherein COSNB is 5-thio-2-nitro-benzoic acid ester."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 27
- (D) OTHER INFORMATION: /label = Xaa  
 /note = "Wherein Xaa is 2-Aminobutyric acid."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly  
 1 5 10 15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val  
 20 25 30

Ile Glu Glu Met Asn Leu Pro Gly  
 35 40

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

(ii) MOLECULE TYPE: peptide

(A) NAME/KEY: Modified-site  
(B) LOCATION: 27  
(D) OTHER INFORMATION: /label = Xaa  
/note = "Wherein Xaa is 2-Aminobutyric acid."

(A) NAME/KEY: Modified-site  
(B) LOCATION: 55  
(D) OTHER INFORMATION: /label = Xaa  
/note = "Wherein Xaa is 2-Aminobutyric acid."

Cys Trp Lys Pro Lys Met Ile Gly Gly Ile Gly Gly Phe Ile Lys Val  
1 5 10 15

Arg Gln Tyr Asp Gln Ile Pro Val Glu Ile Xaa Gly His Lys Ala Ile  
20 25 30

Gly Thr Val Leu Val Gly Pro Thr Pro Val Asn Ile Ile Gly Arg Asn  
35 40 45  
Leu Leu Thr Gln Ile Gly Xaa Thr Leu Asn Phe  
50 55

(2) INFORMATION FOR SEQ ID NO:13:

(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE/TYPE: peptide

[illegible]



- 50 -

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 40
- (D) OTHER INFORMATION: /label = COSBn  
/note = "Wherein COSBn is ??."

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 40
- (D) OTHER INFORMATION: /label = COSBn  
/note = "Wherein COSBn is benzyl thio ester."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly  
1                      5                      10                      15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val  
20                      25                      30

Ile Glu Glu Met Asn Leu Pro Gly  
35                      40

## (2) INFORMATION FOR SEQ ID NO:14:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: peptide

## (ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 40
- (D) OTHER INFORMATION: /label = COSPh  
/note = "Wherein COSPh is phenyl thioester."

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly

09710633-10800

- 51 -

**(2) INFORMATION FOR SEQ ID NO:15:**

(A) LENGTH: 99 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

**(ix) FEATURE:**

(A) NAME/KEY: Modified-site  
(B) LOCATION: 67  
(D) OTHER INFORMATION: //label = Xaa  
/note = "Wherein Xaa is amino butyric acid."

**(ix) FEATURE:**

(A) NAME/KEY: Modified-site  
(B) LOCATION: 95  
(D) OTHER INFORMATION: /label= Xaa  
/note= "Wherein Xaa is 2-Aminobutyric acid."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Pro Gln Ile Thr Leu Trp Lys Arg Pro Leu Val Thr Ile Arg Ile Gly  
1 5 10 15

Gly Gln Leu Lys Glu Ala Leu Leu Asp Thr Gly Ala Asp Asp Thr Val  
20 25 30

Ile Glu Glu Met Asn Leu Pro Gly Cys Trp Lys Pro Lys Met Ile Gly  
35 40 45

Gly Ile Gly Gly Phe Ile Lys Val Arg Gln Tyr Asp Gln Ile Pro Val  
50 55 60

Glu Ile Xaa Gly His Lys Ala Ile Gly Thr Val Leu Val Gly Pro Thr

- 52 -

Leu Asn Phe

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

**(ix) FEATURE:**

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 48  
(D) OTHER INFORMATION: label = COSNB  
/note = "Wherein COSNB is 5-thio-2-nitro benzoic acid ester."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly  
35 40 45

(2) INFORMATION FOR SEQ ID NO:17:

**(i) SEQUENCE CHARACTERISTICS:**

- (A) LENGTH: 62 amino acids  
(B) TYPE: amino acid

**09-06-2008**

- 53 -

- (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear  
(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Cys Ser Ile Gly Gly Asp Ile Phe Ser Asn Arg Glu Gly Lys Leu Pro  
1 5 10 15

Gly Lys Ser Gly Arg Thr Trp Arg Glu Ala Asp Ile Asn Tyr Thr Ser  
20 25 30

Gly Phe Arg Asn Ser Asp Arg Ile Leu Tyr Ser Ser Asp Trp Leu Ile  
35 40 45

Tyr Lys Thr Thr Asp His Tyr Gln Thr Phe Thr Lys Ile Arg  
50 55 60

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site  
(B) LOCATION: 48  
(D) OTHER INFORMATION: /label = COSBn  
/note = "Wherein COSBn is benzyl thio ester."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr  
1 5 10 15

Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala  
20 25 30

008077" 110800

- 54 -

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly  
35 40 45

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 48 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

- (A) NAME/KEY: Modified-site
- (B) LOCATION: 48
- (D) OTHER INFORMATION: /label= COSPh  
/note = "Wherein COSPh is phenyl thio ester."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr  
1 5 10 15  
Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala  
20 25 30  
Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly  
35 40 45

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 110 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

09710633-10800

- 55 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Ala Gln Val Ile Asn Thr Phe Asp Gly Val Ala Asp Tyr Leu Gln Thr  
1 5 10 15

Tyr His Lys Leu Pro Asn Asp Tyr Ile Thr Lys Ser Glu Ala Gln Ala  
20 25 30

Leu Gly Trp Val Ala Ser Lys Gly Asn Leu Ala Asp Val Ala Pro Gly  
35 40 45

Cys Ser Ile Gly Gly Asp Ile Phe Ser Asn Arg Glu Gly Lys Leu Pro  
50 55 60

Gly Lys Ser Gly Arg Thr Trp Arg Glu Ala Asp Ile Asn Tyr Thr Ser  
65 70 75 80

Gly Phe Arg Asn Ser Asp Arg Ile Leu Tyr Ser Ser Asp Trp Leu Ile  
85 90 95

Tyr Lys Thr Thr Asp His Tyr Gln Thr Phe Thr Lys Ile Arg  
100 105 110

09710633-110800